Mutagenesis vol. 27 no. 2 pp. 197–204, 2012 doi:10.1093/mutage/ger061

Ancestral susceptibility to colorectal cancer

Stefanie Huhn[1,*], Barbara Pardini[2], Alessio Naccarati[2], Pavel Vodička[2,3], Kari Hemminki[1,4] and Asta Försti[1,4]

[1]Department of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, 69121 Heidelberg, Germany, [2]Department of Molecular Biology of Cancer, Institute of Experimental Medicine, Academy of Sciences of the Czech Republic, Vídeňská 1083, 142 20 Prague, Czech Republic, [3]Institute of Biology and Medical genetics, 1st Faculty of Medicine, Charles University, Albertov 4, 128 00 Prague 2, Czech Republic and [4]Center for Primary Health Care Research, Lund University, SUS Malmö, at the Clinical Research Center Building 28, Floor 11, Entrance 72, 205 02 Malmö, Sweden.

[*]To whom correspondence should be addressed. Tel: +49 6221 42 1811; Fax: +49 6221 42 1810; Email: s.huhn@dkfz.de

Received on June 20, 2011; revised on August 16, 2011; accepted on August 17, 2011

Every year, approximately 1 million new colorectal cancer (CRC) cases are diagnosed and about half a million people worldwide die due to this cancer. Known differences in CRC incidence rates are mainly attributed to differences in diet and other environmental factors represented, among others, by nutrition-related complex diseases (e.g. obesity and diabetes mellitus type II). Within the last years, it has become evident that environmental risk factors can be complemented by a genetic component when considering the risk of CRC. For example, a number of polymorphisms are known to be associated with an increased risk of obesity and obesity is a risk factor for CRC. Several studies have shown that the 'ancestral-susceptibility model' can be reasonably applied to nutrition-related complex diseases such as obesity. The work in hand shortly discusses whether the ancestral-susceptibility model can also be applied to CRC as a nutrition-related complex disease.

Colorectal cancer epidemiology

Colorectal cancer (CRC) is one of the most common cancers worldwide. The incidence rates for CRC vary among different groups and populations depending on gender, age and country. It is ~1.2 times higher among males than among females and it increases with age (1). It also varies ~25-fold among different populations worldwide. The highest incidence rates can be found in New Zealand, Australia, North America, Western and Northern Europe and more recently in Japan, while lower incidence rates are reported from Africa, Asia and South America (1–4). The differences in CRC incidence rates across the globe are mainly attributed to differences in diet and other environmental factors (3).

CRC—a nutrition-related complex disease

In 1994, Lander and Schork defined complex traits as follows: 'The term ''complex trait'' refers to any phenotype that does not exhibit classic Mendelian recessive or dominant inheritance attributable to a single gene locus' (5). Phenotypes of complex diseases are due to multigenic or multifactorial influences. A disease is called multigenic when more than one gene alteration is required to develop the disease and multifactorial if several genes and environmental factors contribute to risk of developing a disease. The extent of the genetic and environmental contribution to the risk of different diseases is highly variable. In case of multifactorial disease, risk factors frequently show additive effects. An example for a multigenic disease is the Rhesus haemolytic disease, in which two Mendelian loci act together to cause the disease phenotype. On the other hand, the Hirschsprung disease is a well-described example for a multifactorial disease, in which several loci increase the susceptibility to the disease but are not required for the development of the phenotype (6). Multifactorial diseases, affected by both genetic and non-genetic risk factors, are, for example, obesity, diabetes mellitus type II (T2D), Alzheimer's disease and also different types of cancer, including CRC.

In comparison to the hereditary syndromes of CRC, the risk for sporadic CRC can be highly affected by environmental risk factors. Additionally, in the risk of CRC in familial CRC, shared environmental factors play a prominent role. Up to now, six major risk factors have been identified (7–9). Individuals with a body mass index (BMI) $\geq$30 kg/m^2 have a 40% higher risk for CRC than individuals with a BMI $\leq$25 kg/m^2 and the risk of CRC is 20% higher among individuals affected with diabetes compared to unaffected individuals. Additionally, inflammation, alcohol and read meat consumption, cigarette smoking and low physical activity are among the most important risk factors contributing to the risk of CRC (7–10). Next to these most prominent risk factors, vitamin D consumption, calcium intake, fruit and vegetable intake are also under discussion to influence the risk of CRC (7,8).

Within the last years, it has become evident that environmental risk factors themselves can be complemented by a genetic component. In case of complex diseases like obesity and T2D, the evidence is well defined and several polymorphisms are known to be associated with increased risk. But also for risk factors, like cigarette smoking, recent studies have identified genetic variants, influencing smoking behaviour and nicotine dependence (11,12). Other polymorphisms are described influencing the risk of CRC under the effect of environmental factors (9). For example, different alleles of *CYP24A1* affect the CRC risk in association with vitamin D and calcium intake, ultraviolet (UV) exposure and oestrogen replacement therapy (9); a variant of *CCND1* causes an early onset in hereditary non-polyposis cancer but decreases the risk of non-syndromic CRC in combination with oestrogen exposure (13). Considering these interconnections, gene–environment and gene–gene interactions strongly influence the risk of CRC especially involving the complex diseases, such as T2D and obesity, together with malnutrition and supernutrition as key risk factors.

The ancestral-susceptibility model

Nutrition-related complex diseases, such as CRC, obesity and T2D, are prominent diseases of humans living in industrialised societies. This can be attributed to the circumstances that modern societies provide an environment with almost unlimited food supply combined with low energy expenditure that are critical risk factors to CRC, obesity and T2D (2,8,14). The environment provided by industrialised societies may explain some of the high differences in CRC incidence rates across the globe. In an evolutionary relevant timeframe of the human evolution prior to industrialisation, conditions that required protective genetic variants against these diseases were not widely spread. Hence, mechanisms to protect against supernutrition and its health consequences are so far rare either because they have not yet evolved or because they are not yet fixed in human populations (14). However, other features of some of the known risk alleles provide evidence for an ongoing adaptation, and signs of recent adaptation can be found in the human genome (15). A distinct frequency decline from African to European and East Asian populations can be detected for several risk alleles. These frequency differences are getting more pronounced with increasing geographical distance from the ancestral African population (under the assumption of the Out-of-Africa theory). Considering the significant changes of the environmental conditions after leaving Africa ~100 000 years ago, extreme changes in the frequency of the ancestral alleles are an indicator that a selective process may have driven the development of a gene due to a better adaptation to the new environment (14,16,17). Altered requirements in modern and ancestral environments or in geographically different environments may have reduced the advantage of the ancestral allele or even made it harmful (18), while some derived alleles protect against disease and become either neutral or advantageous (19,20). This 'ancestral-susceptibility model' can be reasonably applied to nutrition-related complex diseases (19).

The cultural evolution has changed the human diet fundamentally. According to several studies, the development of agriculture and the domestication of animals have been by far the most important cause of changes in human gene frequencies in the past 10 000 years (14,19–22). Next to the introduction of a variety of new pathogens to the human population, the dietary focus shifted from wild animal protein and fat to domestic animal fat; cereal grains, white flour and cane sugar replaced fresh fruits and vegetables as primary sources of energy. Consequently, the cultural evolution triggered several genetic adaptations, such as the evolution of genetic resistance factors to new crowd-sourced infectious diseases, the evolution of lactose tolerance after ablactation and the evolution of adaptations to the new dietary composition (14,19–22). Food allergies and deficiencies related to animal milk and cereal diets reflect the still incomplete human adaptation to these relatively new food sources (14).

A well-known example for a relatively new adaptation in human evolution is lactose tolerance (23). This trait developed ~5000 years ago outside of Africa, when milk became also available for ablactated children and adults due to the development of agriculture. A new mutation evolved that allowed individuals to digest milk even after ablactation. The ancestral allele of this trait (being lactose intolerant), which had assured that each new born child got enough aliment without rivalry by older siblings, was no longer necessary and even became harmful as milk became staple food. The new more useful gene variant spread fast in those populations that developed agriculture, and its frequency rose. However, the frequency differs considerably among different worldwide populations. Accordingly, the lactose intolerance trait underlies the ancestral-susceptibility model (19).

In case of non-synonymous or coding polymorphisms like lactose tolerance, it is relatively simple to describe the cause and the direction of selective pressure due to the altered function, while in case of non-coding or low-penetrance polymorphisms, complex processes and interactions may cause the final phenotype. Particularly in complex diseases, it should be considered that the detected effect of a low-penetrance polymorphism may affect an intermediate phenotype or that selection had primarily worked on another trait than the analysed one. Among others, possible triggers could be pathogen stress (15,17,24–26) or UV irradiation (15,20,27). Another trigger could be genetic hitchhiking. Genetic hitchhiking describes the genetic linkage between a useful and a harmful variant, whereby the effect of the useful variant is more pronounced (28–32). By this process, selection could have worked on the useful variant, 'accepting' the side effects caused by the linked harmful variant. Additionally, it should be considered that low-penetrance polymorphisms associated with a complex disease can have relatively high frequencies in a population. This might be due to different reasons. In case of a disease that does not influence the reproductive success of an individual, even high-risk alleles for a common disease can spread in a population like neutral polymorphisms. Furthermore, highly variable allele frequencies in different populations might be attributable to processes such as genetic drift or founder effect that occur during the separation from the ancestral population (33).

The expected inheritance model for ancestral-susceptibility traits depends on the trait under consideration. In case of lactose intolerance, the trait is considered recessive since it is a monogenic trait with heterozygote individuals having 98% enzyme activity, which does not influence their ability to digest lactose, whereas homozygote carriers of the ancestral allele are completely unable to digest lactose after infancy. In case of skin colour adaptation or nutrition-related complex diseases, such as CRC, T2D or obesity, which are multifactorial traits, the expected inheritance model is more complicated, most probably including both genetic and environmental factors.

Due to the fact that CRC is a nutrition-related disease that is interlinked with obesity and T2D—both of which have been demonstrated to underlie the ancestral-susceptibility model (19)—it might be possible that CRC itself shows ancestral-susceptibility risk alleles. Furthermore, it might be possible that signs of ongoing selection can be found to alter CRC susceptibility loci.

CTNNBL1 as an ancestral-susceptibility gene (example 1)

One example of a gene in which polymorphisms were found to be associated not only with obesity and high BMI but also with the risk of CRC is the *CTNNBL1* gene [catenin (cadherin-associated protein) b-like 1] (34–37). The study describing the association of *CTNNBL1* with the risk of CRC and fulfilling the characteristics of the ancestral-susceptibility model is shortly reviewed in the following (34).

For the association study, two separate case–control study populations were used. First, a hospital-based study population that originated from the Czech Republic was analysed (38).

This population contained sporadic CRC incident cases and healthy controls. Second, a family-based study population originated from Germany was used (39,40). This population contained case samples with family history of CRC or early onset of CRC and blood donor control samples. From all study participants, DNA has been extracted from peripheral leukocytes by the executive departments in the Czech Republic and Germany. Based on an optimal ratio of directly genotyped versus captured single-nucleotide polymorphisms (SNPs) ($r^2 \geq 0.8$), eight unlinked intronic tagSNPs with a minor allele frequency $\geq$5% were selected (rs6067377, rs2344481, rs238302, rs2281148, rs2235460, rs6067889, rs4811233 and rs6067923) for genotyping. These tagSNPs represent a total of 123 SNPs annotated in the NCBI dbSNP (http://www.ncbi.nlm.nih.gov) encompassing the whole gene. Thus, the SNPs were expected to give information of all common genetic variation within the gene region. The genotype analyses in *CTNNBL1* were performed using allele-specific polymerase chain reaction-based TaqMan assays, designed by Applied Biosystems (TaqMan® SNP Genotyping Assays; Human Pre-Designed Assays Applied Biosystems, Weiterstadt, Germany).

Of the eight selected tagSNPs, three were found to be associated with the risk of CRC. In particular, rs2344481 and rs2235460 showed a modest association of the ancestral allele with an increased risk of CRC in the Czech population and in the analysis of the joint populations (Figure 1). The ancestral allele of rs2281148 showed the opposite effect.

To characterise the gene, phylogenetic networks were applied to the sequence of *CTNNBL1* and the recombination rate and regulatory potential were analysed using UCSC Genome Browser on Human [Mar. 2006 Assembly (hg18) (http://genome.ucsc.edu/) (12,41,42)].

The phylogenetic analysis of the DNA sequences of *CTNNBL1* provided insight about the interspecific conservation of the analysed gene region. The algorithm resulted in highly resolved phylogenetic trees that reflected one of the today's valid phylogenetic trees of animals (43). One tree was based on the complementary DNA (cDNA) sequence of *CTNNBL1*; other trees were based on intronic sequences (Figure 2). The analysis showed that the intron-based phylogenetic trees resembled the cDNA tree in many points. In Figure 2, the phylogenetic tree of intron 3 is shown as an example. The order and species clades were equally well dissolved in the intronic tree as in the cDNA tree. In the case of the primates clade, it fitted the today's valid phylogenetic tree even better than the cDNA tree. Both trees failed in resolving the Laurasiatheria and Euarchontoglires clade (43).

The analysis of the gender-based recombination rate revealed huge differences between males and females. The female recombination rate in *CTNNBL1* was found to be significantly higher than the male one (3.86 times). Additionally, the female recombination rate exceeded that of the gene region and the average chromosomal rate (1.6 and 1.5 times, respectively) (Figure 3B). The regulatory potential in *CTNNBL1* was estimated with 0.06, generally ranging from 0.0 (low) to 0.1 (high).

To characterise the three polymorphisms in *CTNNBL1* that showed a significant association with CRC and to detect possible signs of selection, global allele frequencies, the fixation indexes (F_{ST}) (47), the site-by-site frequency spectrum test of Fay-Wu's H (20,29) and the haplotype test for Standardised integrated haplotype score (iHS) (20,48) were

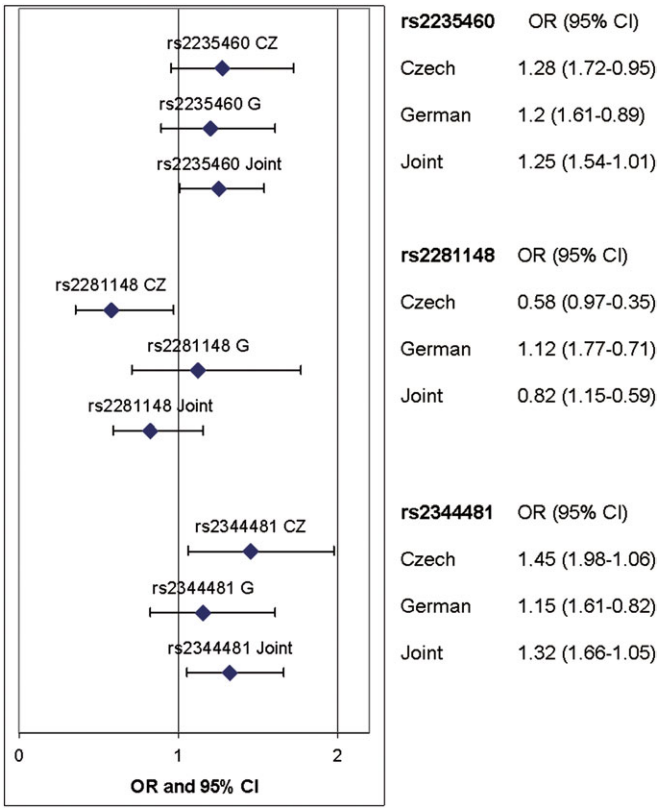

Fig. 1. Data plot of the odds ratios (ORs) and 95% confidence intervals (CIs) of three SNPs in *CTNNBL1* that were found to be significantly associated with the risk of CRC in the Czech cohort. Comparison of the individual results of the Czech and the German cohort (dominant model, unadjusted data) and the results of the joint analysis (dominant model, adjusted for age, gender and nationality). Rhombus OR; bars 95% CI; CZ, Czech samples; G, German samples.

analysed in the Caucasian population (CEU, Utah residents with Northern and Western European ancestry from the CEPH collection), the Sub-Saharan African population (YRI, Yoruba in Ibadan, Nigeria) and the East Asian population (HCB, Han Chinese in Beijing, China, and JPT, Japanese in Tokyo, Japan). The three SNPs showed high worldwide allele frequency differences (>45%; YRI versus CEU versus HCB/JPT); rs606737 and rs2344481 also showed high allele frequency differences between YRI and CEU population (53.5 and 46.5%), respectively, with the ancestral allele frequency changing from the major to the minor allele from YRI to non-African populations. Based on the allele frequency data of the genotyped polymorphisms, the fixation indexes (F_{ST}) and the corresponding probability values were estimated using the Arlequin 3.1 Software (47) (Figure 3A). The F_{ST} is a measure of genetic differences among subpopulations, population differentiation and genetic distance: F_{ST} values >0.25 indicate strong genetic differentiation (i.e. the SNPs might have been targets of selection); F_{ST} values in the range of 0.05–0.1 indicate moderate genetic differentiation (15). F_{ST} P values $\leq$0.05 are generally considered to be evolutionarily significant (49). In *CTNNBL1*, the F_{ST} statistics indicated strong genetic differentiation for the polymorphisms rs2344481 and rs2281148 ($F_{ST} > 0.25$), when the allele frequencies of the African population were compared with the allele frequencies of non-African populations. A moderate genetic differentiation was also found in the comparison of European versus Chinese for rs2344481 and rs2281148. For

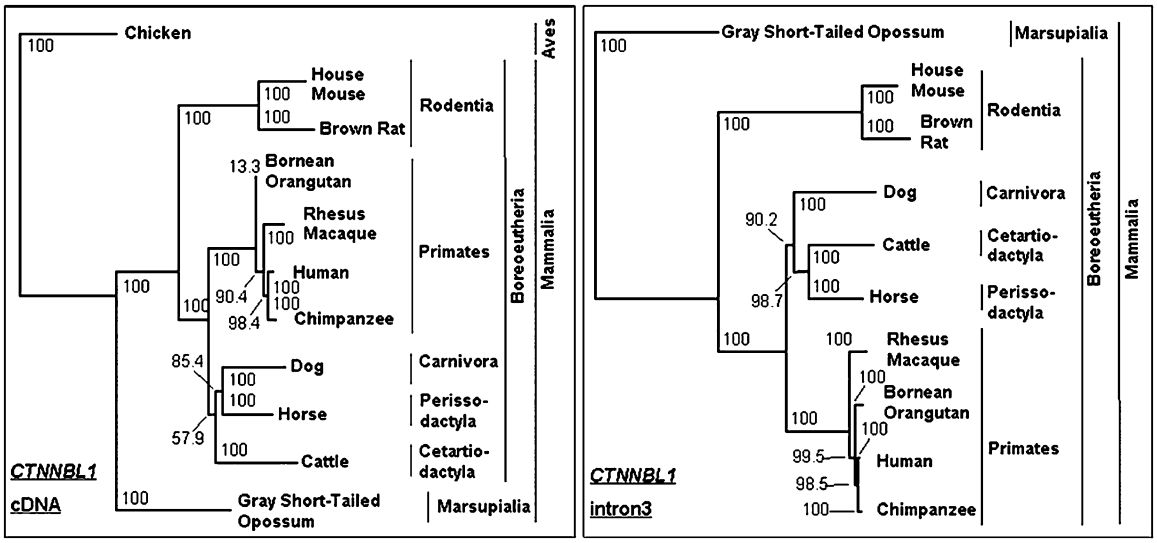

Fig. 2. Phylogenetic trees of the cDNA sequence and intronic DNA (intron 3) sequence of *CTNNBL1* (34). Different taxa are represented by nodes and their evolutionary relationships are represented by branches. Bootstrap values indicate how reliable a clade is. All tested introns gave similar phylogenetic trees as intron 3, exemplarily shown here. Phylogenetic trees of the cDNA sequences and intronic DNA sequences of different species were created with the program SplitsTree V4.10 (44). The sequences of *Homo sapiens sapiens*, *Pan troglodytes*, *Pongo pygmaeus*, *Macaca mulatta*, *Mus musculus*, *Rattus norvegicus*, *Bos taurus*, *Equus freus caballus*, *Canis lupus familiaris* and *Monodelphis domestica* gained from the Ensembl Genome Browser [http://www.ensembl.org/index.html; Hubbard *et al.* (45)] were aligned using MEGA4 software (46). For the cDNA, the sequence of *Gallus gallus domesticus* was additionally used for alignment. The sequence alignments were utilised to calculate phylogenetic consensus trees using the neighbour joining method and to calculate bootstrap values (SplitsTree V4.10) [Huhn *et al.* (34); figure modified and reprinted by permission of Copyright Office e-Century Publishing Corporation; 2011].

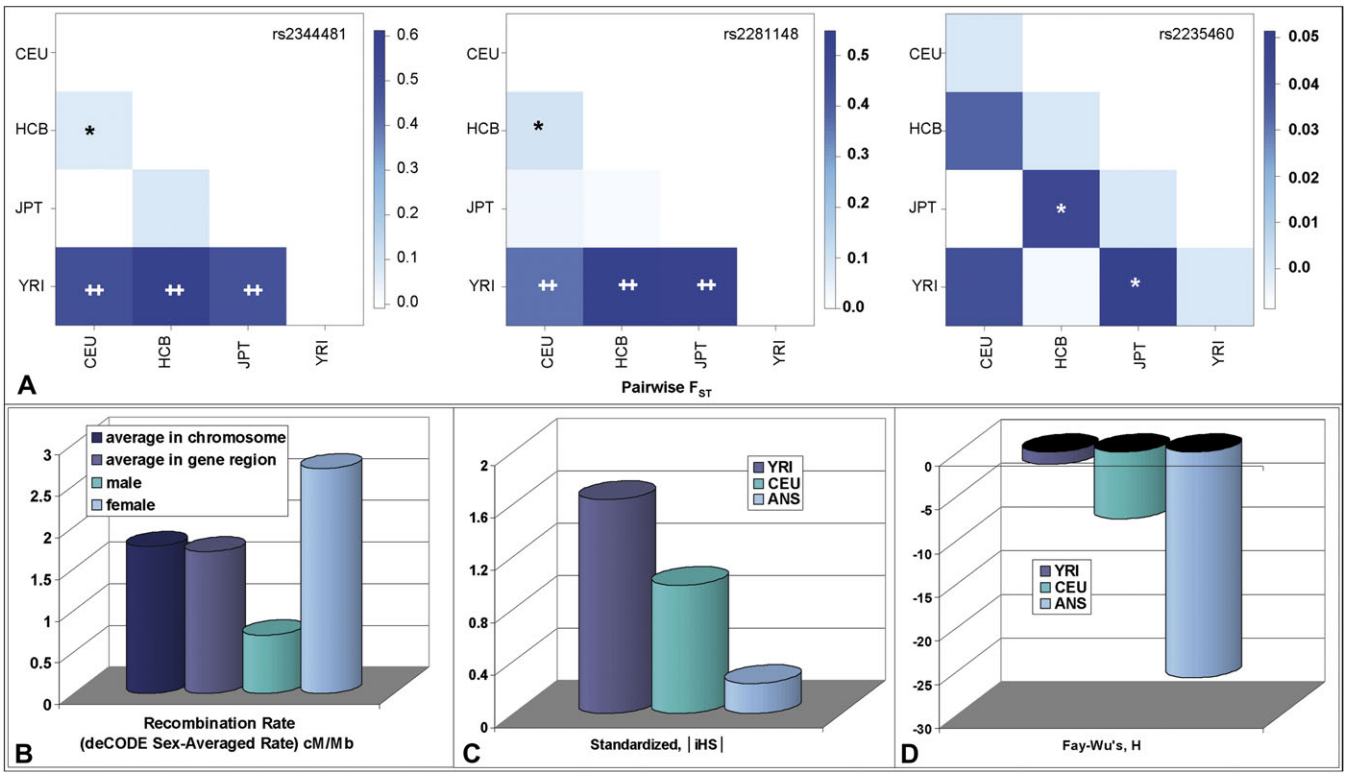

Fig. 3. (**A**) F_{ST} for CRC-associated SNPs in *CTNNBL1* comparing European (CEU), Chinese (HCB), Japanese (JPT) and African (YRI) population. Matrix of pairwise F_{ST}; ‡ indicates strong genetic differentiation and * indicates moderate genetic differentiation at 5% statistical significance. (**B**) Plot of the recombination rate estimated for *CTNNBL1* gene in comparison to the local and average chromosomal rate. (**C**) Plot of the Standardised |iHS| estimate and (**D**) of Fay-Wu's H estimate for rs6020395 in *CTNNBL1*. Rs6020395 was linked to the genotyped SNP rs2344481 ($r^2 > 0.8$). Comparison of the African (YRI), European (CEU) and East Asian (ANS) population.

rs2235460, moderate genetic differentiation (F_{ST} = 0.05) was shown, in a comparison of African and Japanese population as well as Chinese and Japanese population. All estimates that indicated strong or moderate differentiation showed statistical significance at the 5% level, signalling evolutionary significance (15,49).

All estimates of Fay-Wu's H and iHS refer to SNPs that are linked to the genotyped SNPs ($r^2 > 0.8$), due to the fact that direct values of Fay-Wu's H and iHS of the genotyped SNPs were not available in the Haplotter web application (http://haplotter.uchicago.edu/) (20). Strong negative Fay-Wu's H was considered as signatures for a selective sweep (20,29). iHS < -1.5 and > 1.5 give conclusive evidence for natural selection and iHS < -2 or > 2 give evidence for a powerful selection signal (20,48). Signatures of a selective sweep and conclusive evidence for natural selection were found in *CTNNBL1* SNP rs6020395 (Fay-Wu's H = −25.79; |iHS| = 1.63) (Figure 3C and D). This SNP was linked to the genotyped SNP rs2344481 ($r^2 > 0.8$).

These results suggest that the ancestral-susceptibility model may apply to SNPs in *CTNNBL1* and risk of CRC. Additionally, previous studies in *CTNNBL1* indicated an association of polymorphisms within this gene with an increased risk of obesity and increased BMI and fat mass (35–37). The described associations were also conferred by the ancestral alleles of the SNPs analysed in these studies

Table I. Information about genes and SNPs associated with obesity and BMI and possible other nutrition-related traits

Nearest gene	rs No.	Chr.	Position	SNP (A/D)	Risk on A or D	Associated diseases	Function or location	Reference
NEGR1	rs3101336	01p31.01	72.524.023	A/G	D	BMI	*NERG1* ~3 kb upstream	(54)
NEGR1	rs2568958	01p31.01	72.537.954	G/A	D	BMI	*NERG1* ~17 kb upstream	(54)
NEGR1	rs2815752	01p31.01	72.585.278	C/T	D	BMI, obesity, T2D	*NERG1* ~64 kb upstream	(55)
SEC16B	rs10913469	01q25.02	176.180.392	C/T	A	BMI	Intron	(54)
TMEM18	rs2867125	02p24.03	613.077	G/A	A	BMI	*TMEM18* ~45 kb downstream	(54)
TMEM18	rs6548238	02p24.03	625.155	C/T	D	BMI, obesity, T2D	*TMEM18* ~33 kb downstream	(55)
TMEM18	rs4854344	02p24.03	628.394	G/T	D	BMI	*TMEM18* ~30 kb downstream	(54)
TMEM18	rs7561317	02p24.03	635.203	A/G	D	BMI	*TMEM18* ~23 kb downstream	(56)
INSIG2	rs7566605	02q14.01	118.552.745	G/C	D	Obesity	*INSIG2* ~10 kb downstream	(57)
CCDC93	rs11684454	02q14.01	118 479 788	A/G	D	Obesity	Intron	(48)
ETV5; DGKG	rs7647305	03q27.02	187.317.234	T/C	D	BMI	*ETV5* ~8 kb upstream; *DGKG* ~30 kb downstream	(54)
GNPDA2	rs10938397	04p13.00	44.877.534	A/G	D	BMI, obesity, T2D	*GNPDA2* ~454 kb upstream	(55)
PCSK1	rs6235	05q15.00	95.754.904	G/C	D	BMI, obesity, T2D	Missense	(19,58,59)
PCSK1	rs6232	05q15.00	95.777.791	A/G	D	BMI, obesity, T2D	Missense	(58,59)
ADRB2	rs1042714	05q33.01	148.186.916	G/C	A	Obesity, asthma	Missense	(19)
PRL	rs4712652	06p22.03	22.186.844	A/G	A	BMI	*PRL* ~208 kb upstream	(60)
MTMR9	rs2293855	08p23.01	11.215.070	G/A	A	Obesity	Intron	(61)
ADRB3	rs4994	08q12.00	37.943.205	C/T	A	BMI, obesity, T2D	Missense	(19)
GAD2	rs2236418	10p12.01	26.545.752	G/A	A	Obesity	5′ UTR	(19)
GAD2	rs992990	10p12.01	26.226.611	A/C	A	Weight gain, eating behaviour	Intron	(62)
GAD2	rs7908975	10p12.01	26.173.635	C/A	A	Weight gain, eating behaviour	Intron	(62)
PTER	rs10508503	10p13.00	16.340.207	C/T	A	BMI	*PTER* ~179 kb downstream	(60)
PFPK	rs6602024	10p15.02	3.145.487	G/A	D	BMI	Intron	(63)
MTCH2	rs4752856	11p11.02	47.604.868	G/A	D	BMI, obesity, T2D	Intron	(55)
MTCH2	rs10838738	11p11.02	47.619.875	A/G	D	BMI	Intron	(64)
BDNF; BDNFOS	rs4074134	11p14.01	27 604 111	A/G	D	BMI	Intron *BDNFOS*; *BDNF* ~6 kb downstream;	(54)
BDNF; BDNFOS	rs4923461	11p14.01	27 613 736	A/G	A	BMI	Intron *BDNFOS*; *BDNF* ~6 kb downstream;	(54)
BDNF; BDNFOS	rs925946	11p14.01	27 624 028	G/T	D	BMI	Intron *BDNFOS*; *BDNF* ~6 kb downstream;	(54)
BDNF; BDNFOS	rs10501087	11p14.01	27 626 934	T/C	A	BMI	Intron *BDNFOS*; *BDNF* ~6 kb downstream;	(54)

Table 1. *Continued*

Nearest gene	rs No.	Chr.	Position	SNP (A/D)	Risk on A or D	Associated diseases	Function or location	Reference
BDNF; *BDNFOS*	rs6265	11p14.01	27 636 742	G/A	A	BMI	Missense in *BNDF*	(54)
BCDIN3D	rs7138803	12q13.13	48.533.985	G/A	D	BMI	*FAIM2* ~13 kb; *BCDIN3D* ~10 kb	(54)
NRXN3	rs10146997	14q31.01	79.015.165	G/A	A	BMI, waist to hip ratio	Intron	(65)
BBS4	rs7178130	15q24.01	70.765.505	G/A	A	Bardet-Biedl Syndrome, obesity	Before 5′	(58)
ATXN2L; SH2B1	rs8049439	16p11.02	28.745.266	C/T	A	BMI	*ATXN2L* intron; *SH2B1* ~37 kb downstream	(54)
SH2B1	rs4788102	16p11.02	28.781.149	G/A	D	BMI	*SH2B1* ~1.4 kb downstream	(54)
SH2B1	rs7498665	16p11.02	28.790.992	G/A	A	BMI, obesity, T2D	Missense	(55)
FTO	rs6499640	16q12.02	52.327.428	A/G	A	BMI	Intron	(54)
FTO	rs9939973	16q12.02	52.358.319	G/A	D	Obesity	Intron	(66)
FTO	rs1421085	16q12.02	52.358.705	T/C	D	BMI	Intron	(60)
FTO	rs1121980	16q12.02	52.366.998	T/C	A	BMI, obesity, T2D	Intron	(55,66)
FTO	rs8050136	16q12.02	52.374.026	A/C	A	BMI	Intron	(54)
FTO	rs8050136	16q12.02	52.374.026	A/C	A	T2D, BMI	Intron	(48,67,68)
FTO	rs3751812	16q12.02	52.376.211	G/T	D	BMI	Intron	(54)
FTO	rs9939609	16q12.02	52.377.778	A/T	A	BMI	Intron	(67)
FTO	rs9939609	16q12.02	52.378.278	A/T	A	T2D, BMI	Intron	(69)
FTO	rs7190492	16q12.02	52.386.503	G/A	A	BMI	Intron	(54)
FTO	rs9930506	16q12.02	52.388.216	G/A	A	BMI, hip, weight	Intron	(63)
FTO	rs8044769	16q12.02	52.396.886	C/T	A	BMI	Intron	(54)
NPC1	rs1805081	18q11.02	19.394.680	A/G	A	BMI	Missense	(60)
MC4R	rs17782313	18q21.32	56.002.327	T/C	D	BMI	*MC4R* ~187 kb downstream	(60)
MC4R	rs12970134	18q21.32	56.035.980	G/A	D	BMI, T2D	*MC4R* ~53 kb downstream	(54)
KCTD15	rs29941	19q13.11	39.001.622	C/T	A	BMI	*KCTD15* ~6 kb upstream	(54)
KCTD15	rs11084753	19q13.11	39.014.227	G/A	A	BMI, obesity, T2D	*KCTD15* ~17 kb upstream	(55)
CTNNBL1	rs16986921	20q11.23	33117843	T/C	A	BMI, fat mass, obesity	Intron	(35–37)
CTNNBL1	rs6020712	20q11.23	33121934	A/G	A	BMI, fat mass, obesity	Intron	(35–37)
CTNNBL1	rs6013029	20q11.23	33134902	T/G	A	BMI, fat mass, obesity	Intron	(35–37)

Chr, chromosome; A, ancestral allele, D, derived allele.

[rs16986921, rs6020712, rs6013029, odds ratios ~ 1.3]. All three analysed SNPs are in high linkage disequilibrium ($r^2 \geq 0.85$) with the SNP rs2344481 that showed the most significant association with CRC (34). By this, *CTNNBL1* showed shared ancestral susceptibility to CRC and obesity.

CTNNBL1 was suggested as a putative regulator of the canonical Wnt signalling pathway, acting upstream of or in parallel to, β-catenin (34,35). This pathway, if altered or disrupted due to mutations within its components, is known to be related to the development of CRC and other types of cancer (50,51). Furthermore, the Wnt signalling pathway was also shown to be connected to nutrition-related traits, such as normal cholesterol metabolism and glucose-induced insulin secretion and polymorphisms in Wnt signalling components were found to be associated with obesity and T2D (50,52,53).

Other obesity-associated genes following the ancestral-susceptibility model (example 2)

The findings in *CTNNBL1* encouraged the idea to search for other nutrition-associated polymorphisms that also show ancestral susceptibility to a nutrition-related disease and test them for a possible association with CRC. The PubMed database was browsed for genome-wide association studies that focused on obesity or BMI. In 18 publications, 53 polymorphisms were found to be associated with obesity or BMI (Table I).

Of them, 29 SNPs showed ancestral susceptibility and out of them, 9 showed high allele frequency differences (>45%) between African and non-African population as a first indicator of selective pressure. In a first trial, polymorphisms within four of those genes [*MTMR9* (61) (two SNPs), *GAD2* (19,62) (five SNPs), *BBS4* (58) (one SNP) and *NPC1* (60) (one SNP)] were selected to be analysed in an association study in the Czech case–control sample population described in ref. (38). All selected genes encode proteins that have been linked to food intake and obesity, thus being good candidates for CRC susceptibility genes.

The myotubularin-related protein 9 (*MTMR9*) is located at chromosome 8p23–p22. The encoded inactive phosphatase-like molecule interacts with active phosphatases members of the myotubularin-related protein family (MTMR7 and MTMR6). Several original research reports describe a linkage between this locus and obese phenotypes (61,70). MTMR9 expression in the lateral hypothalamic area, the paraventricular- and arcuate-nucleus of the hypothalamus was found to be regulated by diet, especially high-fat diet and might be

involved in appetite regulation (61). The glutamate decarboxylase 2 (*GAD2*) gene encodes for the glutamic acid decarboxylase enzyme (GAD65), which is thought to regulate synaptic release of γ-aminobutyric acid (GABA). The GABA neurotransmitter in the brain impacts most neural functions and is involved in the regulation of food intake (62). The Bardet-Biedl syndrome 4 (BBS4) gene is a member of the BBS gene family. BBS is an autosomal recessive disorder characterised by rod–cone dystrophy, postaxial polydactyly, central obesity, mental retardation, hypogonadism and renal dysfunction (71). BBS proteins were described to be localised to primary cilia (72). The Niemann-Pick disease type C1 (*NPC1*) gene encodes a protein that mediates intracellular cholesterol trafficking. Mutations in *NPC1* and also *NPC2* genes characterise manifest as Niemann-Pick disease, which is a rare, usually fatal, autosomal recessive, lysosomal storage disorder (73).

None of the tested SNPs showed statistically significant associations with CRC (*P* values ranging from 0.21 to 0.79; dominant model; data adjusted for age and gender). However, combining these recent results with those previously described in the Example 1 (*CTNNBL1*), one of five obesity-associated genes showed ancestral susceptibility to CRC and to obesity. Further studies on genes showing ancestral susceptibility to other nutrition-related diseases, such as T2D, cholesterol levels, hypertension and metabolic syndrome, are planned.

The fact that all analysed SNPs were non-coding SNPs does not negate the hypothesis. Several functional non-protein-coding DNA sequences, such as *cis*-regulatory sequences (74) and microRNAs (75), are known to be targets of selection. Polymorphisms associated with multifactorial traits are often located in non-coding sequences, such as in regulatory sequences, 3′ untranslated regions, introns or intergenic sequences of unknown transcriptional status (76). Especially, the new discoveries in the field of epigenetics highlight the role of non-coding DNA sequences in regulatory processes and associations of polymorphisms in these regions with complex diseases (76). It should also be considered that the detected SNP may be linked to a still unknown functional variant as the actual target of selection or that the SNP itself may have a so far unknown regulatory function.

Future perspectives

Due to their complexity, nutrition-related traits and their consequences, intertwining several environmental factors and so far unknown number of genetic variants are great challenges for medical science, disease prevention, therapy and health care policy. A better understanding of the extent of the influence of these factors on the individual risk or on the risk of a population might make it possible to improve prevention, prognosis and treatment. Since susceptibility variants were suggested to show additive effects not only to the risk of cancer but also to the risk of nutrition-related diseases, also variants showing modest effects might be of interest for closer investigation (9). The presented examples suggest that genes found to be associated with one nutrition-related disease might also be promising candidates for other related diseases. The application of additional selection criteria, such as ancestral susceptibility, global allele frequency differences, pathway membership or signatures of selection, might help to narrow down the numerous published polymorphisms and mutations to find the most promising candidates to test in case–control association studies.

Funding

In Germany, the work was supported by the German National Genome Research Network (NGFN-Plus) and EU [HEALTH-F4-2007-200767]. In the Czech Republic, the work was supported by the Czech Republic Grants (CZ:GA CR:GA310/07/1430 and CZ:GA CR:GAP304/10/1286). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Acknowledgements

Conflict of interest statement: None declared.

References

1. Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., Murray, T. and Thun, M. J. (2008) Cancer statistics, 2008. *CA Cancer J. Clin.*, **58**, 71–96.
2. Parkin, D. M., Bray, F., Ferlay, J. and Pisani, P. (2005) Global cancer statistics, 2002. *CA Cancer J. Clin.*, **55**, 74–108.
3. Center, M. M., Jemal, A., Smith, R. A. and Ward, E. (2009) Worldwide variations in colorectal cancer. *CA Cancer J. Clin.*, **56**, 366–378.
4. Ferlay, J., Shin, H. R., Bray, F., Forman, D., Mathers, C. and Parkin, D. M. (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer*, **127**, 2893–2917.
5. Lander, E. S. and Schork, N. J. (1994) Genetic dissection of complex traits. *Science*, **265**, 2037–2048.
6. Strachan, T. and Read, A. P. (1999) *Human Molecular Genetics*. Wiley-Liss, New York, USA.
7. Cunningham, D., Atkin, W., Lenz, H. J., Lynch, H. T., Minsky, B., Nordlinger, B. and Starling, N. (2010) Colorectal cancer. *Lancet*, **375**, 1030–1047.
8. Huxley, R. R., Ansary-Moghaddam, A., Clifton, P., Czernichow, S., Parr, C. L. and Woodward, M. (2009) The impact of dietary and lifestyle risk factors on risk of colorectal cancer: a quantitative overview of the epidemiological evidence. *Int. J. Cancer*, **125**, 171–180.
9. Jasperson, K. W., Tuohy, T. M., Neklason, D. W. and Burt, R. W. (2010) Hereditary and familial colon cancer. *Gastroenterology*, **138**, 2044–2058.
10. Terzic, J., Grivennikov, S., Karin, E. and Karin, M. (2101–2114) Inflammation and colon cancer. *Gastroenterology*, **138**, e2105.
11. Nebert, D. W. and Dalton, T. P. (2006) The role of cytochrome P450 enzymes in endogenous signalling pathways and environmental carcinogenesis. *Nat. Rev. Cancer*, **6**, 947–960.
12. Kong, A., Gudbjartsson, D. F., Sainz, J. *et al.* (2002) A high-resolution recombination map of the human genome. *Nat. Genet.*, **31**, 241–247.
13. Schernhammer, E. S., Tranah, G. J., Giovannucci, E., Chan, A. T., Ma, J., Colditz, G. A., Hunter, D. J., Willett, W. C. and Fuchs, C. S. (2006) Cyclin D1 A870G polymorphism and the risk of colorectal cancer and adenoma. *Br. J. Cancer*, **94**, 928–934.
14. Bellisari, A. (2008) Evolutionary origins of obesity. *Obes. Rev.*, **9**, 165–180.
15. Coop, G., Pickrell, J. K., Novembre, J. *et al.* (2009) The role of geography in human adaptation. *PLoS Genet.*, **5**, e1000500.
16. Novembre, J. and Di Rienzo, A. (2009) Spatial patterns of variation due to natural selection in humans. *Nat. Rev. Genet.*, **10**, 745–755.
17. Sabeti, P. C., Reich, D. E., Higgins, J. M. *et al.* (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature*, **419**, 832–837.
18. Neel, J. V. (1962) Diabetes mellitus: a ''thrifty'' genotype rendered detrimental by ''progress''? *Am. J. Hum. Genet.*, **14**, 353–362.
19. Di Rienzo, A. and Hudson, R. R. (2005) An evolutionary framework for common diseases: the ancestral-susceptibility model. *Trends Genet.*, **21**, 596–601.
20. Voight, B. F., Kudaravalli, S., Wen, X. and Pritchard, J. K. (2006) A map of recent positive selection in the human genome. *PLoS Biol.*, **4**, e72.
21. Diamond, J. (2002) Evolution, consequences and future of plant and animal domestication. *Nature*, **418**, 700–707.
22. Laland, K. N., Odling-Smee, J. and Myles, S. (2010) How culture shaped the human genome: bringing genetics and the human sciences together. *Nat. Rev. Genet.*, **11**, 137–148.
23. Ingram, C. J., Mulcare, C. A., Itan, Y., Thomas, M. G. and Swallow, D. M. (2009) Lactose digestion and the evolutionary genetics of lactase persistence. *Hum. Genet.*, **124**, 579–591.
24. de Silva, E. and Stumpf, M. P. (2004) HIV and the CCR5-Delta32 resistance allele. *FEMS Microbiol. Lett.*, **241**, 1–12.

25. Hamblin, M. T. and Di Rienzo, A. (2000) Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am. J. Hum. Genet.*, **66**, 1669–1679.
26. Tishkoff, S. A., Varkonyi, R., Cahinhinan, N. *et al.* (2001) Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science*, **293**, 455–462.
27. Pickrell, J. K., Coop, G., Novembre, J. *et al.* (2009) Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.*, **19**, 826–837.
28. Barton, N. H. (2000) Genetic hitchhiking. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **355**, 1553–1562.
29. Fay, J. C. and Wu, C. I. (2000) Hitchhiking under positive Darwinian selection. *Genetics*, **155**, 1405–1413.
30. Campbell, M. C. and Tishkoff, S. A. (2008) African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu. Rev. Genomics Hum. Genet.*, **9**, 403–433.
31. Maynard-Smith, J. and Haigh, J. (1974) The hitch-hiking effect of a favourable gene. *Genet. Res.*, **23**, 23–35.
32. Slatkin, M. and Wiehe, T. (1998) Genetic hitch-hiking in a subdivided population. *Genet. Res.*, **71**, 155–160.
33. Slatkin, M. (2004) A population-genetic test of founder effects and implications for Ashkenazi Jewish diseases. *Am. J. Hum. Genet.*, **75**, 282–293.
34. Huhn, S., Ingelfinger, D., Bermejo, J. L. *et al.* (2011) Polymorphisms in CTNNBL1 in relation to colorectal cancer with evolutionary implications. *Int. J. Mol. Epidemiol. Genet.*, **2**, 36–50.
35. Liu, Y. J., Liu, X. G., Wang, L. *et al.* (2008) Genome-wide association scans identified CTNNBL1 as a novel gene for obesity. *Hum. Mol. Genet.*, **17**, 1803–1813.
36. Andreasen, C. H., Mogensen, M. S., Borch-Johnsen, K. *et al.* (2009) Studies of CTNNBL1 and FDFT1 variants and measures of obesity: analyses of quantitative traits and case-control studies in 18,014 Danes. *BMC Med. Genet.*, **10**, 17.
37. Cho, Y. S., Go, M. J., Kim, Y. J. *et al.* (2009) A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat. Genet.*, **41**, 527–534.
38. Pechlivanis, S., Bermejo, J. L., Pardini, B., Naccarati, A., Vodickova, L., Novotny, J., Hemminki, K., Vodicka, P. and Forsti, A. (2009) Genetic variation in adipokine genes and risk of colorectal cancer. *Eur. J. Endocrinol.*, **160**, 933–940.
39. Mangold, E., Pagenstecher, C., Friedl, W. *et al.* (2005) Spectrum and frequencies of mutations in MSH2 and MLH1 identified in 1,721 German families suspected of hereditary nonpolyposis colorectal cancer. *Int. J. Cancer*, **116**, 692–702.
40. Frank, B., Burwinkel, B., Bermejo, J. L. *et al.* (2008) Ten recently identified associations between nsSNPs and colorectal cancer could not be replicated in German families. *Cancer Lett.*, **271**, 153–157.
41. King, D. C., Taylor, J., Elnitski, L., Chiaromonte, F., Miller, W. and Hardison, R. C. (2005) Evaluation of regulatory potential and conservation scores for detecting cis-regulatory modules in aligned mammalian genome sequences. *Genome Res.*, **15**, 1051–1060.
42. Kolbe, D., Taylor, J., Elnitski, L., Eswara, P., Li, J., Miller, W., Hardison, R. and Chiaromonte, F. (2004) Regulatory potential scores from genome-wide three-way alignments of human, mouse, and rat. *Genome Res.*, **14**, 700–707.
43. Murphy, W. J., Eizirik, E., O'Brien, S. J. *et al.* (2001) Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science*, **294**, 2348–2351.
44. Huson, D. H. (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics*, **14**, 68–73.
45. Hubbard, T. J., Aken, B. L., Ayling, S. *et al.* (2009) Ensembl 2009. *Nucleic Acids Res.*, **37**, D690–D697.
46. Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.*, **24**, 1596–1599.
47. Excoffier, L., Laval, G. and Schneider, S. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online*, **1**, 47–50.
48. Southam, L., Soranzo, N., Montgomery, S. B., Frayling, T. M., McCarthy, M. I., Barroso, I. and Zeggini, E. (2009) Is the thrifty genotype hypothesis supported by evidence based on confirmed type 2 diabetes- and obesity-susceptibility variants? *Diabetologia*, **52**, 1846–1851.
49. Wright, S. (1978) *Evolution and the Genetics of Populations: Variability Within and Among Natural Populations*. University of Chicago Press, Chicago, IL, USA.
50. MacDonald, B. T., Tamai, K. and He, X. (2009) Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev. Cell*, **17**, 9–26.
51. Segditsas, S. and Tomlinson, I. (2006) Colorectal cancer and genetic alterations in the Wnt pathway. *Oncogene*, **25**, 7531–7537.
52. Christodoulides, C., Scarda, A., Granzotto, M. *et al.* (2006) WNT10B mutations in human obesity. *Diabetologia*, **49**, 678–684.
53. Jin, T. (2008) The WNT signalling pathway and diabetes mellitus. *Diabetologia*, **51**, 1771–1780.
54. Thorleifsson, G., Walters, G. B., Gudbjartsson, D. F. *et al.* (2009) Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.*, **41**, 18–24.
55. Renström, F., Payne, F., Nordstrom, A., Brito, E. C., Rolandsson, O., Hallmans, G., Barroso, I., Nordstrom, P. and Franks, P. W. (2009) Replication and extension of genome-wide association study results for obesity in 4923 adults from northern Sweden. *Hum. Mol. Genet.*, **18**, 1489–1496.
56. Salonen, J. T., Uimari, P., Aalto, J. M. *et al.* (2007) Type 2 diabetes whole-genome association study in four populations: the DiaGen consortium. *Am. J. Hum. Genet.*, **81**, 338–345.
57. Herbert, A., Gerry, N. P., McQueen, M. B. *et al.* (2006) A common genetic variant is associated with adult and childhood obesity. *Science*, **312**, 279–283.
58. Benzinou, M., Walley, A., Lobbens, S., Charles, M. A., Jouret, B., Fumeron, F., Balkau, B., Meyre, D. and Froguel, P. (2006) Bardet-Biedl syndrome gene variants are associated with both childhood and adult common obesity in French Caucasians. *Diabetes*, **55**, 2876–2882.
59. Kilpeläinen, T. O., Bingham, S. A., Khaw, K. T., Wareham, N. J. and Loos, R. J. (2009) Association of variants in the PCSK1 gene with obesity in the EPIC-Norfolk study. *Hum. Mol. Genet.*, **18**, 3496–3501.
60. Meyre, D., Delplanque, J., Chevre, J. C. *et al.* (2009) Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat. Genet.*, **41**, 157–159.
61. Yanagiya, T., Tanabe, A., Iida, A. *et al.* (2007) Association of single-nucleotide polymorphisms in MTMR9 gene with obesity. *Hum. Mol. Genet.*, **16**, 3017–3026.
62. Choquette, A. C., Lemieux, S., Tremblay, A., Drapeau, V., Bouchard, C., Vohl, M. C. and Perusse, L. (2009) GAD2 gene sequence variations are associated with eating behaviors and weight gain in women from the Quebec family study. *Physiol. Behav.*, **98**, 505–510.
63. Scuteri, A., Sanna, S., Chen, W. M. *et al.* (2007) Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet.*, **3**, e115.
64. Willer, C. J., Sanna, S., Jackson, A. U. *et al.* (2008) Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat. Genet.*, **40**, 161–169.
65. Heard-Costa, N. L., Zillikens, M. C., Monda, K. L. *et al.* (2009) NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium. *PLoS Genet.*, **5**, e1000539.
66. Vogel, C. I., Greene, B., Scherag, A. *et al.* (2009) Non-replication of an association of CTNNBL1 polymorphisms and obesity in a population of Central European ancestry. *BMC Med. Genet.*, **10**, 14.
67. Sandhu, M. S., Weedon, M. N., Fawcett, K. A. *et al.* (2007) Common variants in WFS1 confer risk of type 2 diabetes. *Nat. Genet.*, **39**, 951–953.
68. Scott, L. J., Mohlke, K. L., Bonnycastle, L. L. *et al.* (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*, **316**, 1341–1345.
69. The Wellcome Trust C. C. C. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. *Nature*, **447**, 661–678.
70. Johnson, L., Luke, A., Adeyemo, A. *et al.* (2005) Meta-analysis of five genome-wide linkage studies for body mass index reveals significant evidence for linkage to chromosome 8p. *Int. J. Obes. (Lond.)*, **29**, 413–419.
71. Beales, P. L., Elcioglu, N., Woolf, A. S., Parker, D. and Flinter, F. A. (1999) New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. *J. Med. Genet.*, **36**, 437–446.
72. Shah, A. S., Farmen, S. L., Moninger, T. O. *et al.* (2008) Loss of Bardet-Biedl syndrome proteins alters the morphology and function of motile cilia in airway epithelia. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 3380–3385.
73. Rosenbaum, A. I. and Maxfield, F. R. (2011) Niemann-Pick type C disease: molecular mechanisms and potential therapeutic approaches. *J. Neurochem.*, **116**, 789–795.
74. Wray, G. A. (2007) The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.*, **8**, 206–216.
75. Quach, H., Barreiro, L. B., Laval, G. *et al.* (2009) Signatures of purifying and local positive selection in human miRNAs. *Am. J. Hum. Genet.*, **84**, 316–327.
76. Mattick, J. S. (2009) The genetic signatures of noncoding RNAs. *PLoS Genet.*, **5**, e1000459.